

Face Show

documentation Sept.-Oct. 2023

Gabriel McGee

Alex Both

Darius Shaoul



a chest-up selfie photo of prkrdvs person holding bread in the store, tree in background



a chest-up selfie photo of prkrdvs person wearing a black shirt, holding bread in the store, tree in background

the angle and quality of the photo, including how much of the person you can see

the token name + the word person.

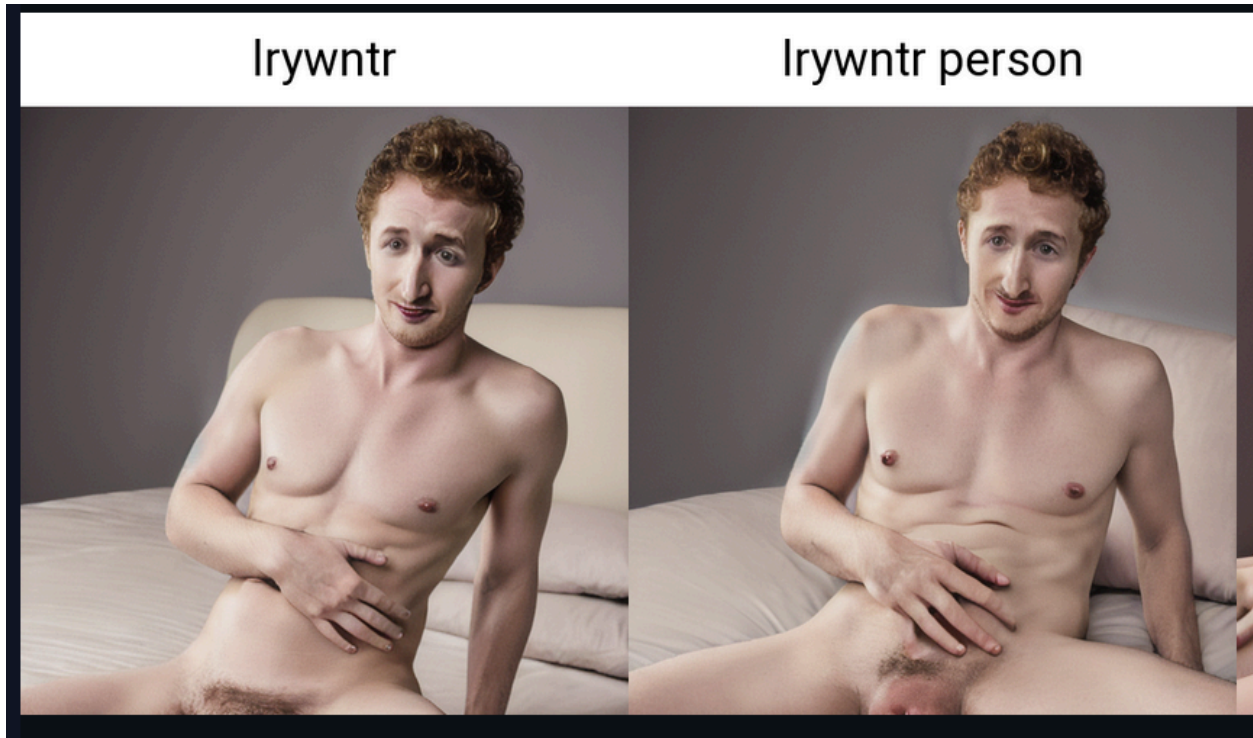
characteristics of the person, include anything that's unique about the photo and the figure which shouldn't be learned in the training (e.g. parker does not always have a black shirt and isn't always in the store). Be specific about anything in the photo. I didn't tag parker's glasses because he always has glasses on.

1. train ai on everyone's face
 - a. get 3-5 selfies from each RSVP'd person
 - b. train textual inversion over the week w/ each person,
 - c. take photos of people in the gallery to train it on the location
 - d. create a template for the text-to-image generations
2. create text to speech
 - a. make a list of conversation prompts for inserting speech-to-text input
 - b. have microphone do speech to text and record samples from the party
 - c. create a framework to send text from microphone into opensource LLM
 - d. use the transcribed conversation snippets to create text-to-image prompt and chatbot prompt at the same time
 - e. project photo outputs with text rendered below, use tts to read out the dialogue.
 - i. link photo with text output under the name of the speaker, send to one of two screens somehow
 - ii.
 - iii.
 - iv.
 - v.

Process notebook

notes on prompts:

use the keyword and the weight like this: (lrywnttr :1). adding person unnecessarily makes it more "person-like",



regional prompter



candid photo of 3people at a hipster party BREAK

(nnsmn :1) BREAK
(gbrlmcg :1) BREAK
(drsshl :1)

Steps: 20, Sampler: DPM++ 2M Karras, CFG scale: 7, Seed: 4191126256, Size: 1024x512, Model hash: 79b5166ee6, Model: last-flipped2cond-ep400-gs05500, RP Active: True, RP Divide mode: Matrix, RP Matrix submode: Columns, RP Mask submode: Mask, RP Prompt submode: Prompt, RP Calc Mode: Attention, RP Ratios: "1,1,1", RP Base Ratios: 0.2, RP Use Base: True, RP Use Common: False, RP Use Ncommon: False, RP Change AND: False, RP LoRA Neg Te Ratios: 0, RP LoRA Neg U Ratios: 0, RP threshold: 0.1, RP LoRA Stop Step: 0, RP LoRA Hires Stop Step: 0, RP Flip: False, Version: v1.6.0

MULTI FACE ADETAILER

<https://github.com/Bing-su/adetailer/wiki/Advanced>

<https://github.com/Bing-su/adetailer/discussions/204>

normal prompt:

a

adetailer prompt:

(nnsnmn :1), 1girl [SEP] (lxbth :1), 1girl [SEP] (gbrlmcg :1), 1boy [SEP], (drsshl :1), 1boy

LORA model training

- follow github tutorial for runpod (may need diff tut for alienware)
 - <https://github.com/FurkanGozukara/Stable-Diffusion/blob/main/Tutorials/How-To-Install-Kohya-LoRA-Web-UI-On-RunPod.md#auto-installer-for-runpod>
- download regularization images:
 - wget
<https://github.com/FurkanGozukara/Stable-Diffusion/blob/main/Tutorials/How-To-Install-Kohya-LoRA-Web-UI-On-RunPod.md#auto-installer-for-runpod>
- Testing 3 images of alex, 1024x1024, max res 512x512
- alextest1
 - using sdxl produces massive checkpoints for lora. tested w/ sdxl base
 - woman at a party as class identifier
- alextest2
 - testing using all default except no xformers and network size of 256.
 - woman as class identifier
 - base model:
/workspace/stable-diffusion-webui/models/Stable-diffusion/Realistic_Vision_V5.1.safetensors
- alextest3
 - testing using all default except no xformers and network size of 8.
 - woman as class identifier
- alextest4
 - network size 8
 - woman at a party with white walls, size 8 network
- alextest5
 - network size 8
 - repeats 20 (instead of 40)
 - xformers on
 - 1024x1024
 - woman in a white room class prompt
- alextest6 (20230917) alex's computer
 - same as alextest4
 - network size 8
 - new reg images from party
- alextest7 gabe's computer
 - new reg images
 - "woman in white room"
- alextwst8 alex's computer
 - new reg images
 - 20 epochs
 - 60 network size
- alextest9 darius computer
 - network rank 8
 - repeats:60
- alextest10 darius computer
 - batch size 2
 - constant lr scheduler
 -

Current test is with iranian dude's recommendation of a network size of 256. this may be way the lora models are big but probably makes good results. another option from [reddit](#) says to make `lora_rank` smaller

NEW TECHNIQUE

alexndarius test 1

use automatic1111 dreambooth training

supposedly, if no .txt is found, it reads the name of the file as the instance token.

use [filewords] in the instance *token*

use person for class token

use full body photo of [filewords] person for instance prompt

use full body photo of person for class prompt

profit?

realized issue. the filenames are not the same. all the tokens will be different. you need to reinstalling microsoft malware or you will to be losing all of investment portolio to hacker sir

->might need to create .txt files containing a unique instance token(1 per subject) for each image.

alexndarius test 2

make a txt file corresponding to each photo like this picture

->

use the same instance token in each person's txt files (in this example im just removing the vowels in people's names)

use [filewords] in the instance *token*

use person for class token

use full body photo of [filewords] person for instance prompt

use full body photo of person for class prompt

alexndarius test 3_3

all of the same but using sd1.5 base model and v1x model and class images... (basically the known functioning shit)

results:

best results with learning rate of 2e-6 at 1500 steps (over 8000 was bad...)

trying with RV using model type 1.x(rv was trained on 1.5) and ema+unfreeze

multi-subject lora tests on Kohya

first test:

repeats dont matter as long as all are around 100, (its for balancing multiple images seperately)

standard lora with 32 network size is pretty good.

defaults are pretty good

repeats balancing should be used if some folders contain more images than others

multi-subject and place test (dreambooth realistic vision 512) test 1

captioned place photos, left subjects with 'ohxw person'

multi-subject sdxl w/ 1300 reg images

good. only works at 1024x1024. using exactly secourses json. repeats =86 per img folder

faces were not learned well enough.

name

 alex0.jpg

 alex1.jpg


 alex2.jpg

 alex3.jpg


 alex4.jpg


 alex5.jpg


 alex6.jpg

 darius0.jpg

 darius1.jpg


 darius2.jpg


 darius3.jpg


 darius4.jpg

 darius5.jpg


 darius6.txt


 darius6.jpg


 darius5.txt


 darius5.jpg


 darius4.txt


 darius4.jpg


 darius3.txt

 darius3.jpg


 darius2.txt


 darius2.jpg

 darius1.txt

 darius1.jpg

 darius0.txt

 darius0.jpg

 alex6.txt

  **darius3.txt**

drssh1

multi-subject Realistic vision xl 2 w/ 500 reg images

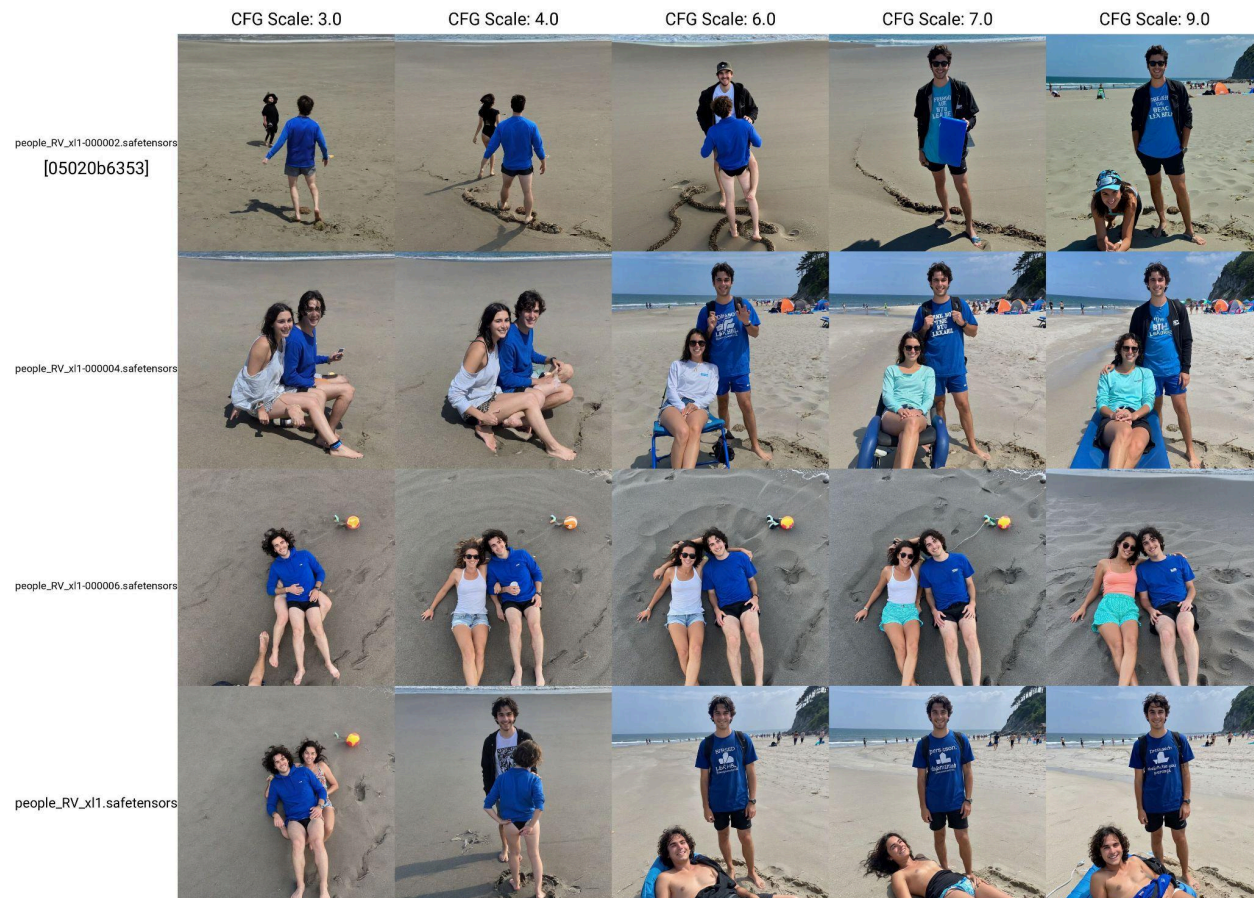
the idea: less reg images = better face memorization, lower step count. doing 33 repeats per folder. max steps 7920

results

soooo the images are all beautiful - however the faces get mixed quite easily

more steps = more defined faces. This works way better than other models for some reason.

created a basic grid:



rvxl-lora big boy test

settings:

```
accelerate launch --num_cpu_threads_per_process=2 -- ./sd_xl_train_network.py"
--pretrained_model_name_or_path="C:\Users\student\Downloads\sd.webui\models\Stable-diffusion\rvxl-training2_RealVisXL_V2.0.safetensors"
--train_data_dir="C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\img"
--resolution="1024,1024" --output_dir="C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\model"
--logging_dir="C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\log" --network_alpha="1"
--save_model_as=safetensors --network_module=networks.lora --text_encoder_lr=0.0004
--unet_lr=0.0004 --network_dim=32 --output_name="rvxl-celeb-noreg"
--lr_scheduler_num_cycles="10" --no_half_vae --full_bf16 --learning_rate="0.0004"
--lr_scheduler="constant" --train_batch_size="1" --max_train_steps="4950"
--save_every_n_epochs="1" --mixed_precision="bf16" --save_precision="bf16"
--caption_extension=".txt" --cache_latents --cache_latents_to_disk --optimizer_type="AdamW"
--optimizer_args scale_parameter=False relative_step=False warmup_init=False
--max_data_loader_n_workers="0" --bucket_reso_steps=64 --gradient_checkpointing --xformers
--bucket_no_upscale --noise_offset=0.0

prepare tokenizers
Using DreamBooth method.
prepare images.
found directory C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\img\33_alannamasterson person contains 5 image files
found directory C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\img\33_greydamon person contains 5 image files
found directory C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\img\33_matvairo person contains 5 image files
495 train images with repeating.
0 reg images.
no regularization images / 正则化画像が見つかりませんでした
[Dataset 0]
batch_size: 1
resolution: (1024, 1024)
enable_bucket: False

[Subset 0 of Dataset 0]
image_dir: C:\Users\student\Desktop\kohya\kohya_ss\rvxl-lora\img\33_alannamasterson person"
image_count: 5
num_repeats: 33
shuffle_caption: False
keep_tokens: 0
```


[illegible]

CFG Scale: 7.0



results boii

random T1 test (probs retarded)
following this guys ideas
<https://www.reddit.com/r/StableD>

EveryDream2 testing

test1. testing using 5 photos each
= 5. next test i will try custom bat

test 2. batch size 1 with custom batch
batch_size=5, so i changed it to

- result:
 - underfit in com
 - Batch size of 1
 - custom batchir

test3. gonna continue to try custom
flipped images

```
%run train.py --config train.j
--resumeckpt "SG161222/Realis
--project_name "flipped-custom
--data_root "input" \
--max_epochs 400 \
--sample_steps 150 \
--save_every_n_epochs 100 \
--lr 1.2e-6 \
--lr_scheduler constant \
--save_full_precision \
--batch_size 5
```

- result:
 - underfit even a
 - custom batchir

good group photo including (1g

test 4: lower cond dropout, no cu

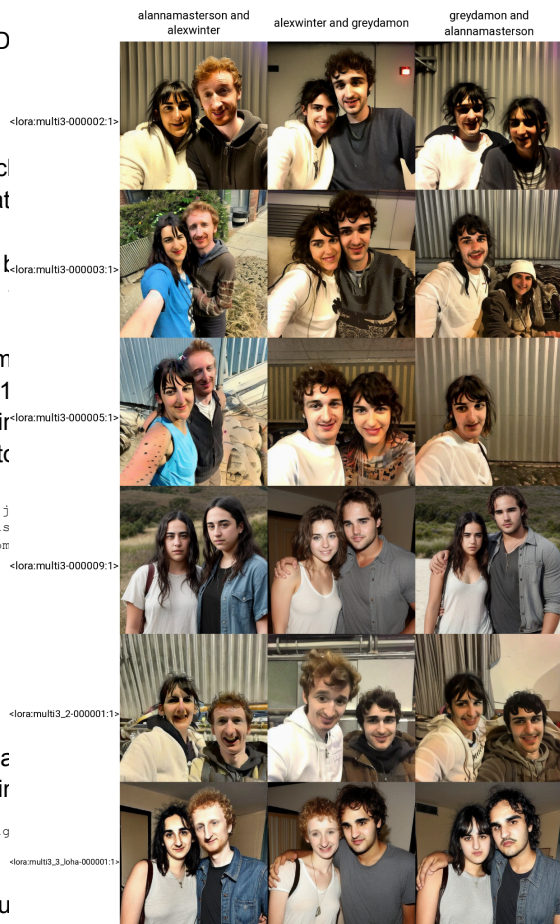
```
%run train.py --config train.json \
--resumeckpt "SG161222/Realistic_Vision_V5.1_noVAE" \
--project_name "flipped2cond" \
--data_root "input" \
--max_epochs 400 \
--sample_steps 150 \
--save_every_n_epochs 100 \
--lr 1.2e-6 \
--lr_scheduler constant \
--save_full_precision \
--batch_size 5 \
--cond_dropout 0.02
```

test5 same but w/ 0.07 balancing in the gt folder using multiply.txt, and batch size of 2

```
%run train.py --config train.json \
--resumeckpt "SG161222/Realistic_Vision_V5.1_noVAE" \
--project_name "flipped2condbalanced" \
--data_root "input" \
--max_epochs 400 \
--sample_steps 150 \
--save_every_n_epochs 100 \
--lr 1.2e-6 \
--lr_scheduler constant \
--save_full_precision \
--batch_size 2 \
--cond_dropout 0.02
```

Test 4 vs test 5

verdict: test 4 @ 400epochs is best.

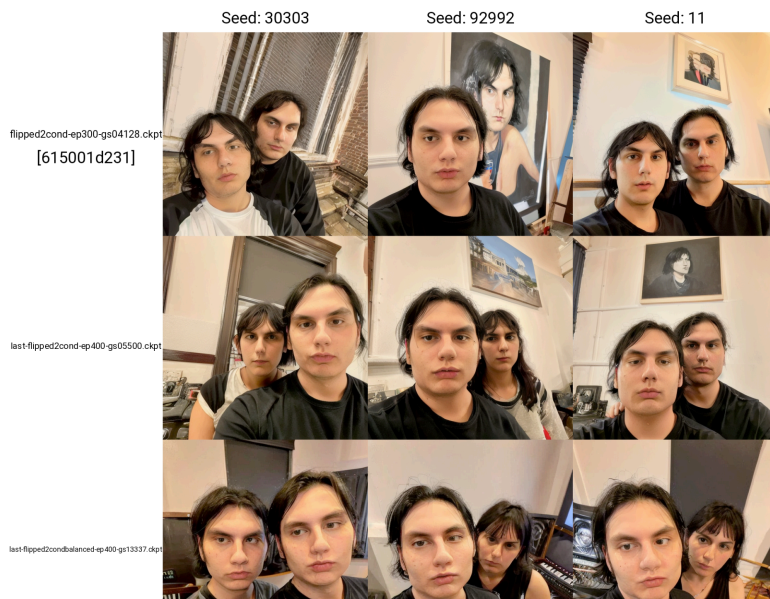


multiple_subjects/

xt. only modification is batch size

out 5 images per bucket with
size.

ze=5, doubling the dataset w/



current best settings

normal prompt including multiple characters, hires fix, then use adetailer with [sep] tags in adetailer prompt, mediapipe_full_face.pt,

main prompt

amazing photo of 4people at the beach, (nnsnmn :1.2) 1girl, and (lxbth :1.2) 1girl, and (gbrlmcg :1.2) 1boy, and (drsshl :1.2) 1boy

adetailer prompt

(nnsnmn :1.2) 1girl [SEP] (lxbth :1.2) 1girl [SEP] (gbrlmcg :1.2) 1boy [SEP], (drsshl :1.2) 1boy



Seed: 13123

Seed: 232

flipped-custom-batch-ep300-gs04109.ckpt

[22a4abe3db]



last-flipped-custom-batch-ep400-gs05489.ckpt

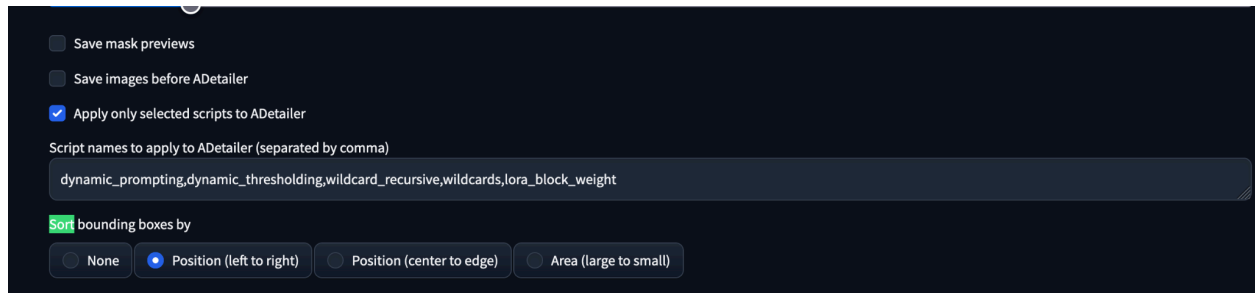
[15511cf88b]



last-noval-ep250-gs02236.ckpt



use left-right bounding boxes if you want control of the order the faces are created



SDXL dreambooth 2

epoch 15 might be best, lowest loss around there

using new patreon settings, no reg

IT SUCKS

back to everydream2 (tentative idea to chain it in into a higher resolution at more steps?)

added 500 more GT images, using multiply.txt of 0.02 ($0.02 \times 700 = 14$ per batch).

```
%run train.py --config train.json \
--resume_ckpt "SG161222/Realistic_Vision_V5.1_noVAE" \
--project_name "flipped2cond" \
--data_root "input" \
--max_epochs 400 \
--sample_steps 150 \
--save_every_n_epochs 25 \
--lr 1.5e-6 \
--lr_scheduler constant \
--save_full_precision \
--batch_size 5 \
--cond_dropout 0.02
```

^ decent results; will prob be good enough w/ adetailer.

testing at 764 w/ a lower learning rate


```
%run train.py --config train.json \
--resume_ckpt "SG161222/Realistic_Vision_V5.1_noVAE" \
--project_name "flippedcond764" \
--data_root "input" \
--max_epochs 400 \
--sample_steps 150 \
--save_every_n_epochs 50 \
--lr 0.8e-6 \
--lr_scheduler constant \
--save_full_precision \
--batch_size 5 \
--cond_dropout 0.02 \
--resolution 768
```





oct 17 running main model + batched sulkroom using batch_id only in sulk. 200 gt images

```
%run train.py --config train.json \
--resume_ckpt "SG161222/Realistic_Vision_V5.1_noVAE" \
--project_name "flipped2cond" \
--data_root "input" \
--max_epochs 350 \
--sample_steps 150 \
--save_every_n_epochs 50 \
--lr 1.5e-6 \
--lr_scheduler constant \
--save_full_precision \
```

```
--batch_size 10 \
--cond_dropout 0.02
```

replicate / vicuna-13b

 PUBLIC A large language model that's been fine-tuned on ChatGPT interactions

 176.2K runs  GitHub  Paper  License

[▶ Demo](#) [🔗 API](#) [📄 Examples](#) [🕒 Versions \(6282abe6\)](#)

Input

T prompt

subtitles for alex both starting with: "thats literally so dumb why would you say that"

Prompt to send to Llama.

max_length

500

Maximum number of tokens to generate. A word is generally 2-3 tokens (minimum: 1)

temperature

SYSTEM

write a monologue script for a gossiping college student at a party trying to get attention, written in 2019 on tiktok

Output

Alex:im sorry, i was trying to be funny
Alex2: its fine, comedy can be hit or miss
Alex: yeah, sometimes my jokes fall flat
Alex2: no worries, just be yourself
Alex: thanks, i appreciate that
Alex2: of course, you're a great person
Alex: aw, thanks


Generated in 6.30 seconds

USER write a short snippet of dialogue spoken by a ridiculous college student at an art school party. it begins with: "thats literally so dumb why would you say that..."

ASSISTANT (Record button clicked)
"Oh, my gosh! That's literally so dumb, why would you even say that? Okay sunscreen enthusiasts, listen up! Now Jeremy from Design 101 has ditched his uptight country club look and is officially dialing up his edgy quotient. He dyed his hair fuchsia last night, yes FUCH-SIAA! Exactly like Fiona's hair! Related or suspect?"


(Turns camera to self with a fake dramatic gasp)

"And hey, did you guys know? Sarah's vertical garden that she can't stop posting about, which by the way, got 500 more likes? Bought from an online store! Ikea has suddenly beamed into our world, disrupted our noble efforts and is

 Add message

Submit



 Give us feedback

Mode

 Chat

Model

gpt-4

Temperature 1.15

Maximum length 150

Stop sequences
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

 API and Playground requests will not be used to train our

SYSTEM

write a short casual party-gossip snippet from the perspective of an informal, 17 year old jock pervert, the script begins with the sentence: "thats literally so dumb why would you say that..." . |

obviously craves and deserves?

Critics might say that I'm out of line, that I'm overstepping boundaries. But, isn't it just the natural order? We're beings of sensory pleasure, driven by our primal instincts. So why should I stifle these passions, ignore the beautiful panorama of desire that's spread out before me?

They can prattle on about decency and decorum

ASSISTANT

until the cows come home. But I'll tell you this, there's a difference between appreciating beauty and being a creep. I've never forced myself onto anyone, never crossed those lines.

But man, I can't deny the pull...the sweet, sweet magnetism. It's the game of glances, the dance of smiles, the teasing and the chasing. Even now, I can feel the eyes of a couple of pretty things on me, their shy smiles promising more.

Call me what you want, but I'm just an average guy navigating the maddening maze of teenage desires. Ain't nothing wrong in that, right?"

vii.

Chat GPT system qualifiers:

SYSTEM

write a FIRST-PERSON monologue script for a gossiping young people at an art party, written in 2019 on tiktok, retarded teens chatting loosely, do not set the stage, only write dialogue. the conversation topics are outrageous and touch on eclectic topics emotionally, including gossip and intrigue within the party.

USER continue this conversation: "have you talked to leroy lately?"

+ Add message

Submit ↻

Mode

Chat

Model

gpt-4

Temperature 1.21

Maximum length 150

Stop sequences

Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0.36

API and Playground requests will not be used to train our models. [Learn more](#)

Hey chat, I would like you to compose a FIRST-PERSON script of gossiping young people at an art party, written in 2023 on tiktok, retarded teens chatting loosely. Do not set the stage, only write dialogue, the conversation topics are outrageous and touch on eclectic topics emotionally, including gossip and intrigue within the party. Below I have created 2 main lists. The first list is of Names I would like you to use exclusively for names of the characters in the script you generate. The second list is of Chicago Neighborhoods, Venues, and Bars. You can use these locations as part of the conversation, but you DO NOT HAVE TO. Maybe 0-2 locations per script.

Names: Darius, Alex, Gabriel, Leroy, Allegra, Parker, Nell, Ava, Henry, Max B., Anna, Lucas

Neighborhoods: The loop, Gold Coast, Old Town, South Loop, Pilsen, Chinatown Lakeview, Wrigleyville, Wicker Park, Bucktown, Ukrainian Village, Bridgeport, Little village, Wigga Park, Logan, East Pilsen, River North, Gold Coast, Albany Park

Venues: The Bottle, Clipper, Podlasie, Smart bar, Archer, Laura, Sleeping Village, Cafe Moustachio, The Hideout, Thalia Hall, DADS, Co-Prosperity Sphere, Not Not, Charleston

Bars: Skylark, Rainbo, Inner town, Cole's, The Continental, Richards, Punch House,

Things can also take place at School or campus. Thanks Chat.

sulk embedding instructions:

put this at the end of the prompt:

[person in the room of sulk3:1]

download the embedding from

https://drive.google.com/file/d/1oO_g3OmZrSiD5_EXE88VT8d6K9niewM6/view?usp=share_link

put the sulk3.bin into the "embeddings" folder in stable-diffusion-webui

Random example

((4 people at a party)), group photo, (gracerequejo :1.2) 1girl, and (Cassandrafitzgerald:1) 1boy, and (johncook:1.2) 1boy, and (yigitural:1.2) 1boy, beautiful, dancing, soft focus, close up, dark lighting, selfie, 8k photo

(gracerequejo :1) [SEP] (Cassandrafitzgerald:1) [SEP] (johncook:1) [SEP] (yigitural:1)

RANDOMNESS PROMPTS:

RANDOMNESS:

Prompt: UHD photo of (___people__:1.1), (___people__:1.1), (___people__:1.1), (___people__:1.1) at *blah blah*

ADetailer: (___people__:1.1) [SEP] (___people__:1.1) [SEP] (___people__:1.1) [SEP] (___people__:1.1)

(___people__:1) [SEP] (___people__:1) [SEP] (___people__:1)

MIXING REGIONAL PROMPTING AND DYNAMIC PROMPTING AND THE SULK EMBEDDING

Prompt:

UHD flash photo of 3people, (candid party photography) [person in the room of sulk3-1000] BREAK

(___people__:1.1)BREAK

(___people__:1.1)BREAK

(___people__:1.1)

best settings ever

prompt

3people faces photo (candid party flash photography) multiple people in the white walls room [sulk3-1000] BREAK

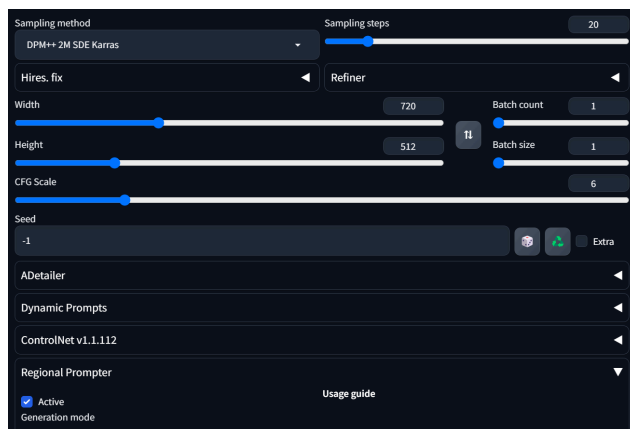
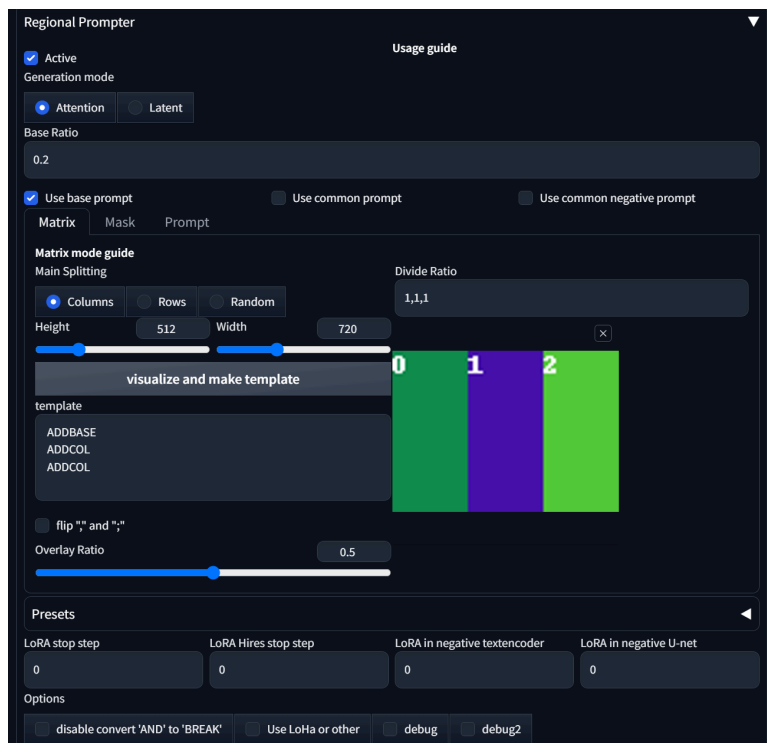
3 people, (__people__) BREAK

3 people, (__people__) BREAK

3 people, (__people__)

negative:

(deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, anime, mutated hands and fingers:1.4), (deformed, distorted, disfigured:1.3), poorly drawn, bad anatomy, wrong anatomy, extra limb, missing limb, floating limbs, disconnected limbs, mutation, mutated, ugly, disgusting, amputation,



3people faces photo (candid party flash photography) multiple people in the white walls room [sulk3] BREAK

3 people, __angle__ (__people__) __activites__ BREAK

3 people, __angle__ (__people__) __activites__ BREAK

3 people, __angle__ (__people__) __activites__

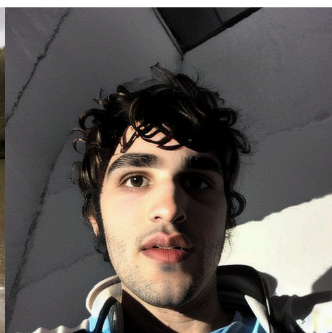
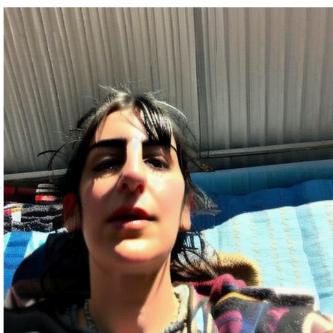


alannamasterson

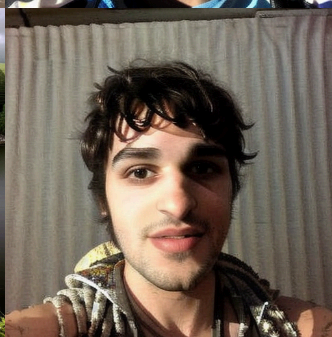
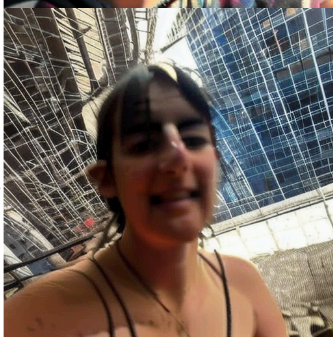
alexwinter

greysdamon

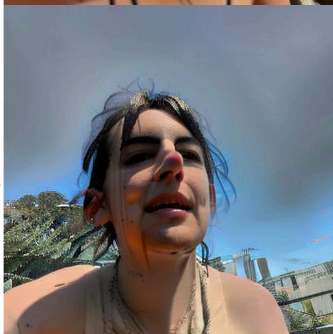
<lora:multi3-000002:1>



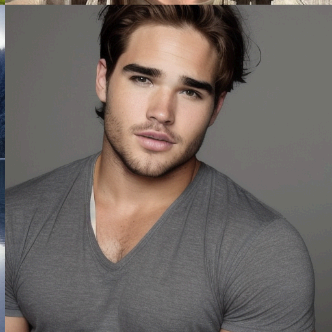
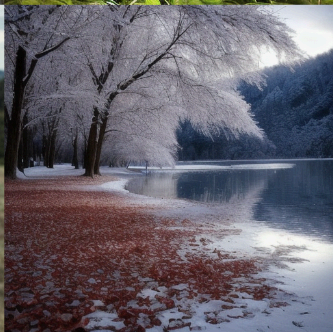
<lora:multi3-000004:1>



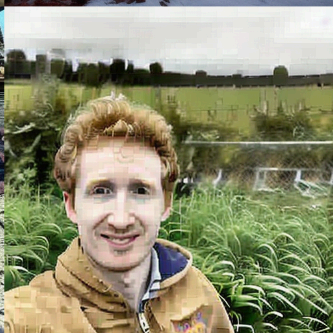
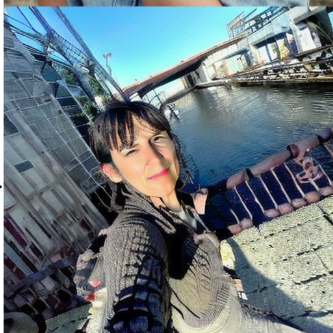
<lora:multi3-000006:1>



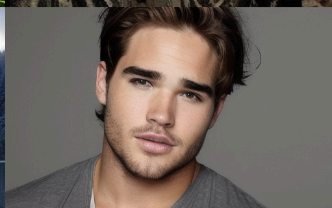
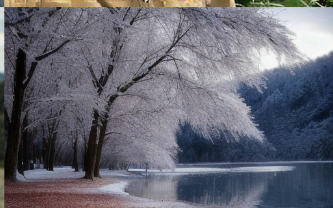
<lora:multi3-000008:1>



<lora:multi3_2-000002:1>



<lora:multi3_3_laha-000002:1>

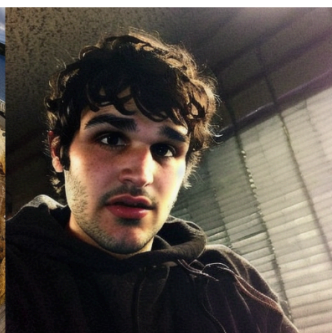


alannamasterson

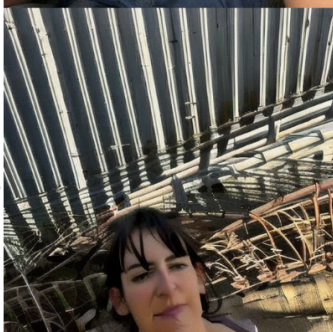
alexwinter

greyclamon

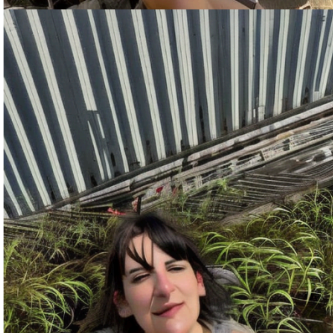
<lora:multi3-000001:1>



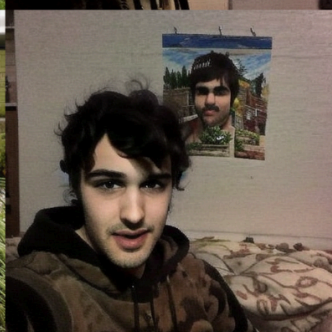
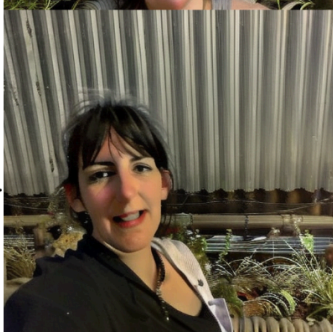
<lora:multi3-000004:1>



<lora:multi3:1>



<lora:multi3_2-000001:1>



<lora:multi3_3_laha-000001:1>

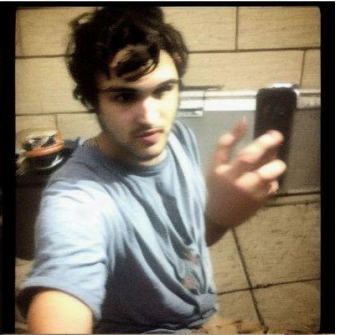


alannamasterson person

alexwinter person

greyclamon person

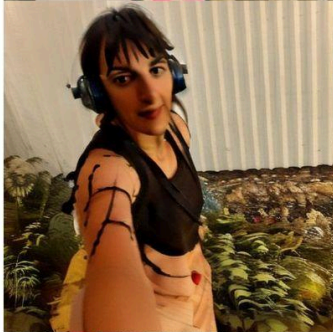
<lora:multi3-000001:1>



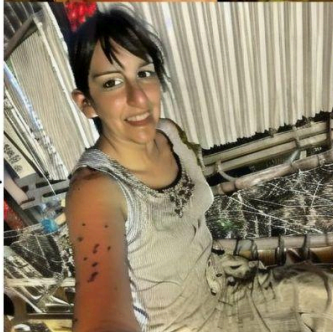
<lora:multi3-000004:1>



<lora:multi3:1>



<lora:multi3_2-000001:1>



<lora:multi3_3_Joha-000001:1>

